On Convergence of Proportionate-Type Normalized Least Mean Square Algorithms

Rajib Lochan Das and Mrityunjoy Chakraborty, *Senior Member, IEEE*

***Abstract*—In this paper, a new convergence analysis is presented for a well-known sparse adaptive filter family, namely, the proportionate-type normalized least mean square (PtNLMS) algorithms, where, unlike all the existing approaches, no assumption of whiteness is made on the input. The analysis relies on a "transform" domain based model of the PtNLMS algorithms and brings out certain new convergence features not reported earlier. In particular, it establishes the universality of the steady-state excess mean square error formula derived earlier under white input assumption. In addition, it brings out a new relation between the mean square deviation of each tap weight and the corresponding gain factor used in the PtNLMS algorithm.**

***Index Terms*—Excess mean square error (EMSE), PtNLMS algorithm, sparse adaptive filters, stability of EMSE.**

I. Introduction

In practice, one often encounters systems that have a sparse impulse response with the degree of sparseness varying with time and context. Typical examples include network echo paths [1], acoustic echo channels between microphone and loudspeaker in handheld mobile telephony [2], channels in high-definition television [3], wireless multipath channels in cellular communication [4], acoustic channels in underwater communication [5], etc. During the last decade or so, several sparse adaptive filters have been proposed to identify such systems, which offer improved convergence performance as compared with sparsity agnostic adaptive algorithms such as the LMS or normalized LMS (NLMS) [6] by making appropriate use of the system sparseness. A detailed account of these developments is given in [7].

Of all the sparse adaptive filters proposed so far, the proportionate-type NLMS (PtNLMS) algorithms [8]–[10] constitute by far the most popular class of algorithms. The PtNLMS algorithms try to identify and track a sparse system modeled by a $L \times 1$ finite-impulse response filter coefficient vector $\mathbf{w}_{\rm opt}$, which takes an input $x(n)$ and produces an output $y_d(n) = \mathbf{w}_{\rm opt}^T\mathbf{x}(n) + v(n)$, where $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-L+1)]^T$ is the input data vector, and $v(n)$ is an observation noise that is usually modeled as zero mean, Gaussian, independent identically distributed, and independent of $x(m)$ for any n, m. The vector $\mathbf{w}_{\rm opt}$ is identified by updating a filter coefficient vector $\mathbf{w}_P(n) = [w_{P,0}(n), w_{P,1}(n), \ldots, w_{P,L-1}(n)]^T$ as [11]

$$\mathbf{w}_P(n+1) = \mathbf{w}_P(n) + \frac{\mu\mathbf{G}(n)\mathbf{x}(n)e(n)}{\mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n) + \delta_P} \quad (1)$$

where $\mathbf{G}(n)$ is a diagonal gain matrix, with $g_i(n) \equiv [\mathbf{G}(n)]_{i,i}$ satisfying $\sum_{i=0}^{L-1} g_i(n) = 1$ $(0 < g_i(n) < 1)$, δ_P is a very small positive constant deployed to avoid division by zero, μ is the step size, and $e(n) = y_d(n) - \mathbf{w_P}^T(n)\mathbf{x}(n)$ is the filter output error. In all the PtNLMS algorithms, the gain factor $g_i(n)$ is made proportional to $|w_{P,i}(n)|$ in some way or other, which means that the effective step size $\mu g_i(n)$ assumes large values for active taps and small values for inactive taps, leading to a combination of higher convergence speed and lesser steady-state excess mean square error (EMSE).

A mean square error (MSE) analysis of the PtNLMS algorithms is, however, an exceedingly difficult task in comparison with, for example, the NLMS algorithm due to the presence of the gain matrix $\mathbf{G}(n)$ in the numerator and the denominator of the update term in (1), where $\mathbf{G}(n)$ itself depends on $\mathbf{w}_P(n)$. Conventional approaches to such analysis [8], [12]–[14] try to get around this problem by assuming the input to be white, which is, however, not very realistic in certain system identification contexts (e.g., echo cancellation). In this brief, we present a new convergence analysis of the PtNLMS algorithms, where we evaluate the steady-state EMSE and conditions for stability of the EMSE *without imposing any whiteness assumption on the input*. The analysis relies on a "transform" domain model of the PtNLMS algorithms [15], [16] and brings out certain new convergence features not reported earlier. In particular, it establishes the universality of the steady-state EMSE formula derived earlier in [8] under white input assumption. In addition, it brings out a new relation between the mean square deviation (MSD) of each tap weight and the corresponding gain factor $g_i(n)$.

II. Proposed Convergence Analysis

A. Transform Domain Model of the PtNLMS Algorithms

The transform domain model of the PtNLMS algorithms, used indirectly in [15] and [16] earlier, is shown in Fig. 1, where $\mathbf{G}^{1/2}(n)$ is a transform matrix, with $[\mathbf{G}^{1/2}(n)]_{i,i} = g_i^{1/2}(n)$, $i = 0, 1, \ldots, L-1$; and $\mathbf{s}(n)$ and $\mathbf{w}_N(n)$ are the so-called "transformed" input and filter coefficient vectors, respectively, with $\mathbf{s}(n) = \mathbf{G}^{1/2}(n)\mathbf{x}(n)$ and $\mathbf{w}_N(n) = [\mathbf{G}^{1/2}(n)]^{-1}\mathbf{w}_P(n)$.

Manuscript received July 13, 2014; accepted November 29, 2014. Date of publication December 25, 2014; date of current version April 23, 2015. This brief was recommended by Associate Editor G. J. Dolecek.

The authors are with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India (e-mail: rajib.das.iit@gmail.com; mrityun@ece.iitkgp.ernet.in).

Color versions of one or more of the figures in this brief are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSII.2014.2386261

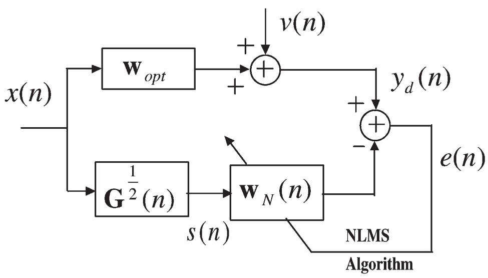

Fig. 1. NLMS-based transform domain model of the PtNLMS algorithms.

It is easy to check that $\mathbf{w}_N^T(n)\mathbf{s}(n) = \mathbf{w}_P^T(n)\mathbf{x}(n) \equiv \mathbf{y}(n)$ (say), i.e., the filter $\mathbf{w}_N(n)$ with input vector $\mathbf{s}(n)$ produces the same output $y(n)$ produced by $\mathbf{w}_P(n)$ with input vector $\mathbf{x}(n)$. To compute $\mathbf{G}^{1/2}(n+1)$ and $\mathbf{w}_N(n+1)$, the filter $\mathbf{w}_N(n)$ is first updated by the NLMS weight update recursion to a weight vector $\mathbf{w}'_N(n+1)$ as $\mathbf{w}'_N(n+1) = \mathbf{w}_N(n) + \frac{\mu\mathbf{s}(n)e(n)}{\mathbf{s}^T(n)\mathbf{s}(n)+\delta_P}$. From (1), $\mathbf{w}_P(n+1)$ is then computed as $\mathbf{w}_P(n+1) = \mathbf{G}^{1/2}(n)\mathbf{w}'_N(n+1)$. The matrix $\mathbf{G}(n+1)$ follows from $\mathbf{w}_P(n+1)$, following its definition, and $\mathbf{w}_N(n+1)$ is then evaluated as $\mathbf{w}_N(n+1) = [\mathbf{G}^{1/2}(n+1)]^{-1}\mathbf{w}_P(n+1)$. Note that $\mathbf{w}_N(n+1) = \mathbf{G}^{-(1/2)}(n+1)\mathbf{w}_P(n+1) = \mathbf{G}^{-(1/2)}(n+1)\mathbf{G}^{1/2}(n)\mathbf{w}'_N(n+1)$, which means that $[\mathbf{w}_N(n+1)]_i = [g_i(n)/g_i(n+1)]^{1/2}[\mathbf{w}'_N(n+1)]_i$, $i = 0, 1, \ldots, L-1$. Since $\sum_{i=0}^{L-1} g_i(n) = 1$ and $0 < g_i(n) < 1$, $i = 0, 1, \ldots, L-1$, it is reasonable to expect that $g_i(n)$ does not change significantly from index n to index $(n+1)$ [particularly near convergence and/or for large order filters], and thus, we can make the approximation $[g_i(n)]^{1/2}[\mathbf{w}'_N(n+1)]_i \approx [g_i(n+1)]^{1/2}[\mathbf{w}'_N(n+1)]_i$, which implies $\mathbf{w}'_N(n+1) = \mathbf{w}_N(n+1)$.

B. Proposed Analysis

For our analysis here, we approximate δ_P by zero in (1) as δ_P is a very very small constant. Defining the weight error vector $\tilde{\mathbf{w}}_P(n) = \mathbf{w}_{\text{opt}} - \mathbf{w}_P(n)$ and the transformed weight error vector $\tilde{\mathbf{w}}_N(n) = \mathbf{G}^{-(1/2)}(n)\tilde{\mathbf{w}}_P(n) \equiv \mathbf{G}^{-(1/2)}(n)\mathbf{w}_{\text{opt}} - \mathbf{w}_N(n)$, it is easy to check that $e(n) = \tilde{\mathbf{w}}_N^T(n)\mathbf{s}(n) + v(n)$. Let $\mathbf{e}_{s,i}(n)$ be the ith normalized eigenvector of the correlation matrix $\mathbf{S}(n) = E[\mathbf{s}(n)\mathbf{s}^T(n)]$ corresponding to the eigenvalue $\lambda_{s,i}(n)$, which means that $\mathbf{S}(n)$ can be factorized as $\mathbf{S}(n) = \mathbf{E}_s(n)\mathbf{D}_s(n)\mathbf{E}_s^T(n)$, where $\mathbf{E}_s(n) = [\mathbf{e}_{s,0}(n), \mathbf{e}_{s,1}(n), \ldots, \mathbf{e}_{s,L-1}(n)]$, and $\mathbf{D}_s(n) = \text{diag}(\lambda_{s,i}(n))$. In addition, define the weight error correlation matrix $\mathbf{K}_s(n) = E(\tilde{\mathbf{w}}_N(n)\tilde{\mathbf{w}}_N^T(n))$ and its transformed version $\mathbf{\Lambda}_s(n) = \mathbf{E}_s^T(n)\mathbf{K}_s(n)\mathbf{E}_s(n)$. Then, assuming statistical independence between $\mathbf{w}_N(n)$ and $\mathbf{s}(n)$ (which follows from the commonly assumed independence between $\mathbf{w}_P(n)$ and $\mathbf{x}(n)$ and the aforementioned assumption that the elements of $\mathbf{G}(n)$ have marginal variance, particularly in steady state) and following the standard procedure to evaluate the output MSE in an adaptive filter [6], [17], it is straightforward to show that $\xi(n) = \xi_0 + \text{Tr}(\mathbf{K}_s(n) \times \mathbf{S}(n)) = \xi_0 + \text{Tr}(\mathbf{\Lambda}_s(n)\mathbf{D}_s(n))$, leading to

$$\xi(n) = \xi_0 + \sum_{i=0}^{L-1} \lambda_{s,i}(n)\tilde{\lambda}_{s,i}(n) \tag{2}$$

where $\tilde{\lambda}_{s,i}(n) = [\mathbf{\Lambda}_s(n)]_{i,i}$, and ξ_0 is the minimum MSE obtainable, which, in this case, is given by the variance σ_v^2 of the observation noise $v(n)$.

As seen from (2), evaluation of the steady-state MSE $\xi(\infty)$ requires an analysis of the evolution of $\tilde{\lambda}_{s,i}(n)$ with time, which is quite a complicated task. To simplify the analysis, we employ here a scheme of angular discretization of a continuous valued random vector proposed first by Slock [18] and used recently in [19] in the context of affine-projection-based adaptive filters. As per this, given a zero-mean random vector $\mathbf{x}$ with correlation matrix $\mathbf{R} = E[\mathbf{x}\mathbf{x}^T]$, it is assumed that $\mathbf{x}$ can assume only one of the $2L$ directions, given by $\pm\mathbf{e}_i$, $i = 0, 1, \ldots, L-1$, where $\mathbf{e}_i$ is the ith normalized eigenvector of $\mathbf{R}$ corresponding to the eigenvalue λ_i. In particular, $\mathbf{x}$ is modeled as $\mathbf{x} = s\,r\,\mathbf{v}$, where $\mathbf{v} \in \{\mathbf{e}_i | i = 0, 1, \ldots, L-1\}$, with probability of $\mathbf{v} = \mathbf{e}_i$ given by p_i, $r = \|\mathbf{x}\|$, i.e., r has the same distribution as that of $\|\mathbf{x}\|$, and $s \in \{1, -1\}$, with a probability of $s = \pm 1$ given by $P(s = \pm 1) = 1/2$. Furthermore, the three elements s, r, and $\mathbf{v}$ are assumed to be mutually independent. Note that, as s is zero mean, $E[s\,r\,\mathbf{v}] = \mathbf{0}$, and thus, $E[\mathbf{x}] = \mathbf{0}$ is satisfied trivially. To satisfy $E[\mathbf{x}\mathbf{x}^T] = \mathbf{R}$, the discrete probability p_i is taken as $p_i = \lambda_i/\text{Tr}\,[\mathbf{R}]$, which leads to $E[\mathbf{x}\mathbf{x}^T] = E(s^2\,r^2\,\mathbf{v}\mathbf{v}^T) = E(r^2)\,E(\mathbf{v}\mathbf{v}^T) = \text{Tr}\,[\mathbf{R}]\sum_{i=0}^{L-1} p_i\mathbf{e}_i\mathbf{e}_i^T = \sum_{i=0}^{L-1}\lambda_i\mathbf{e}_i\mathbf{e}_i^T = \mathbf{R}$ (note that the above choice of p_i satisfies the two basic requirements of discrete probability, namely, $p_i \geq 0$ and $\sum_{i=0}^{L-1} p_i = 1$). In addition, note that, if θ_i is the angle between $\mathbf{x}$ and $\mathbf{e}_i$, then $\cos(\theta_i) = \mathbf{x}^T\mathbf{e}_i/\|\mathbf{x}\|$ and $E[\cos^2(\theta_i)] \approx \lambda_i/\text{Tr}\,[\mathbf{R}] = p_i$, which means that p_i provides a measure of how far $\mathbf{x}$ is (angularly) from $\mathbf{e}_i$ on an average.

Applying the above to the present context, $\mathbf{s}(n)$ can be written as

$$\mathbf{s}(n) = s_s(n)\,r_s(n)\,\mathbf{v}_s(n) \tag{3}$$

where $s_s(n) \in \{+1, -1\}$, with $P(s_s(n) = \pm 1) = 1/2$, $r_s(n) = \|\mathbf{s}(n)\|$, and $\mathbf{v}_s(n) \in \{\mathbf{e}_{s,i}(n) | i = 0, 1, \ldots, L-1\}$, with $P(\mathbf{v}_s(n) = \mathbf{e}_{s,i}(n)) = \lambda_{s,i}(n)/\text{Tr}(\mathbf{S}(n))$, where, as before, the three elements $s_s(n)$, $r_s(n)$, and $\mathbf{v}_s(n)$ are assumed to be mutually independent. It may be mentioned here that time dependence of the eigenvectors $\mathbf{e}_{s,i}(n)$ and the eigenvalues $\lambda_{s,i}(n)$ makes the analysis using the model (3) a lot more difficult than all other cases that use the same model while also assuming the input to be WSS, which renders the preceding eigenvectors and eigenvalues time invariant (note that $s(n)$ cannot be WSS even when $x(n)$ is because of its dependence on $\mathbf{G}(n)$).

Using (3), it is nevertheless possible to prove the following.

Theorem 1: With a zero-mean Gaussian input $x(n)$ of variance $E(x^2(n)) = \sigma_x^2$, the PtNLMS algorithms produce stable MSE performance if the step size μ satisfies $0 < \mu < 2$, and the corresponding steady-state MSE is given by

$$\xi(\infty) = \xi_0 + \frac{\xi_0\sigma_x^2\mu}{\hat{r}_s^2(\infty)(2-\mu)}$$

where ξ_0 is the minimum MSE obtainable ($\equiv E(v^2(n)) = \sigma_v^2$), and $1/\hat{r}_s^2(\infty) = \lim_{n\to\infty} 1/\hat{r}_s^2(n)$, where $1/\hat{r}_s^2(n) = E[1/r_s^2(n)]$, and $r_s^2(n) = \mathbf{x}^T(n)\mathbf{G}(n)\mathbf{x}(n)$.

Proof: The proof is given in the Appendix. ■

From Theorem 1, the steady-state EMSE of the PtNLMS algorithms is given by $\xi_{\text{excess}} = \xi_0 \sigma_x^2 \mu / (\hat{r}_s^2(\infty)(2-\mu))$. Several other observations also follow from Theorem 1, which are embodied in the following corollaries.

Corollary 1: For long filters, the steady-state MSEs of PtNLMS algorithms are almost independent of the gain matrix, i.e., all PtNLMS algorithms produce approximately the same steady-state MSE, which is given by

$$\xi(\infty) = \xi_0 + \xi_0 \frac{\mu}{(2-\mu)}. \tag{4}$$

Proof: When the filter length is large, the variance of $r_s^2(n) = \|\mathbf{s}(n)\|^2$ is very small, and $1/\hat{r}_s^2(n)$ can be approximated as $1/\hat{r}_s^2(n) = E[1/r_s^2(n)] \approx 1/E(r_s^2(n))$. Since $E(r_s^2(n)) = \mathrm{Tr}(\mathbf{S}(n)) = \mathrm{Tr}(E[\mathbf{G}^{1/2}(n)\mathbf{R}\mathbf{G}^{1/2}(n)]) = \sigma_x^2$ (as $\sum_{i=0}^{L-1} g_i(n) = 1$), the result follows trivially from Theorem 1. ■

Corollary 1 reveals that all PtNLMS algorithms with a constant step size μ attain the same steady-state MSE irrespective of whether the input is white or colored, which is also the same MSE obtained under NLMS. Incidentally, the MSE expression given in (4) is the same as that obtained in [8] *under a white input assumption*. In other words, Corollary 1 establishes the universality of the $\xi(\infty)$ expression derived in [8].

Corollary 2: For a white input, the steady-state MSD for the ith tap in a PtNLMS algorithm, $i = 0, 1, \ldots, L-1$, is linearly proportional to the steady-state mean value of the corresponding gain element, as given by

$$\lim_{n\to\infty} E\left(\tilde{w}_{P,i}^2(n)\right) = \frac{\mu \xi_0}{\sigma_x^2 (2-\mu)} \bar{g}_i(\infty) \tag{5}$$

where $\bar{g}_i(\infty) = \lim_{n\to\infty}[E(\mathbf{G}(n))]_{i,i}$.

Proof: For a white input $x(n)$ with variance σ_x^2, the autocorrelation matrix $\mathbf{R} = \sigma_x^2 \mathbf{I}$, and thus, we have $\mathbf{S}(n) = E(\mathbf{G}^{1/2}(n)\mathbf{R}\mathbf{G}^{1/2}(n)) = \sigma_x^2 E(\mathbf{G}(n))$: a diagonal matrix, which means that the eigenvalues of $\mathbf{S}(n)$ are given by $\lambda_{s,i}(n) = \sigma_x^2 \bar{g}_i(n)$, $i = 0, 1, \ldots, L-1$, and the corresponding eigenvectors are given by $\mathbf{e}_{s,i}(n) = [0, \ldots, 0, 1, 0, \ldots, 0]^T$, where "1" occurs in the ith position. Clearly, $\mathbf{E}(n) = \mathbf{I}$, and thus, $\tilde{\lambda}_{s,i}(n) = [\mathbf{\Lambda}_s(n)]_{i,i} = [\mathbf{E}^T(n)\mathbf{K}_s(n)\mathbf{E}(n)]_{i,i} = [\mathbf{K}_s(n)]_{i,i} = [E(\tilde{\mathbf{w}}_N(n)\tilde{\mathbf{w}}_N^T(n))]_{i,i} = E(\tilde{w}_{N,i}^2(n))$. Now, the expression for $\tilde{\lambda}_{s,i}(\infty)$ has been worked out in the Appendix and is given by (A.5). From this and the above, for white input, we then have $\lim_{n\to\infty} E(\tilde{w}_{N,i}^2(n)) = \tilde{\lambda}_{s,i}(\infty) = \xi_0\mu/(\hat{r}_s^2(\infty)(2-\mu)) \approx \xi_0\mu/(\sigma_x^2(2-\mu))$. Recalling $\tilde{\mathbf{w}}_N(n) = \mathbf{G}^{-(1/2)}(n)\tilde{\mathbf{w}}_P(n)$, we can then write $\lim_{n\to\infty} E(\tilde{w}_{P,i}^2(n)) = \lim_{n\to\infty} E(g_i(n)\tilde{w}_{N,i}^2(n)) \approx \lim_{n\to\infty} E(g_i(n))E(\tilde{w}_{N,i}^2(n)) = \frac{\xi_0\mu}{\sigma_x^2(2-\mu)}\bar{g}_i(\infty)$, where, as before, we assume that $g_i(n)$ has marginal variance in the steady state and thus is largely uncorrelated with $\tilde{w}_{N,i}^2(n)$ as $n \to \infty$. ■

The MSD expression given in (5) is a refinement of an earlier result obtained in [12], which expresses the MSD as $\lim_{n\to\infty} E(\tilde{w}_{P,i}^2(n)) = \frac{\xi_0\mu\bar{g}_i(\infty)}{\sigma_x^2(2-\mu\bar{g}_i(\infty))}$. Simulation results given in the next section confirm that the MSD given by (5) is much closer to reality than the one proposed in [12].

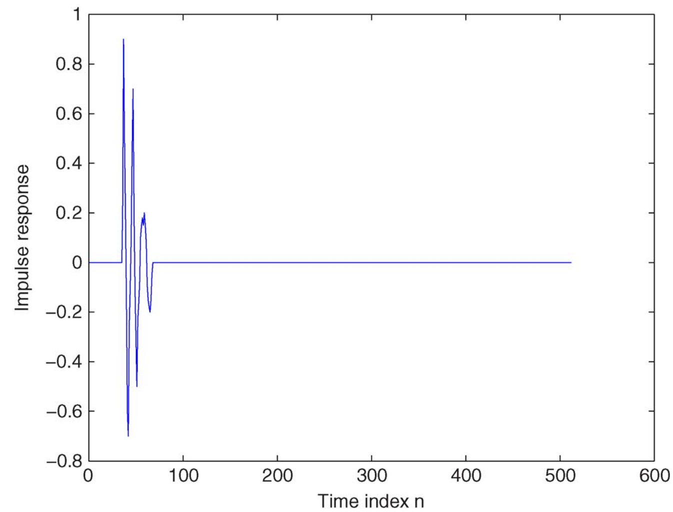

Fig. 2. Impulse response of the sparse network echo path chosen for identification.

III. Simulation Studies

For simulation, we considered a sparse network echo path as the unknown sparse system with a typical impulse response, as shown in Fig. 2, having altogether 32 active taps out of a total of 512 taps, located within the 36th to 67th tap positions. The observation noise $v(n)$ at the system output was taken to be zero-mean white Gaussian, with $\sigma_v^2 = 10^{-3}$. Simulation was carried out for both colored and white input $x(n)$, with $\sigma_x^2 = 1$. The colored input was generated by driving an AR(1) model with a zero-mean unity variance white process $u(n)$ as $x(n) = rx(n-1) + \sqrt{1-r^2}u(n)$ (r was chosen to be 0.8). For simulation, two PtNLMS algorithms, namely, the PNLMS and the IPNLMS, were considered along with the NLMS for filters having a total of 512 taps. For all the algorithms, the step size μ and the regularization parameter (to avoid division by zero) were taken to be 1 and 0.01, respectively. Furthermore, for the PNLMS [8], we set $\rho = 0.01$ and $\delta = 0.001$, whereas for the IPNLMS [9], we fixed $\alpha = -0.5$. The simulation was carried out for a total of 20 000 iterations, and the corresponding results are shown in Fig. 3 by plotting the learning curves, i.e., the EMSE $\xi(n)$ against the iteration index n, obtained by averaging $(\tilde{\mathbf{w}}_P^T(n)\mathbf{x}(n))^2$ over 400 experiments. The figure confirms that, for white and colored input, the PNLMS, IPNLMS, and NLMS algorithms eventually produce the same steady-state EMSE, which is seen to be -30 dB for the given case. Furthermore, this also matches our conjecture (i.e., Theorem 1) that $\xi_{\text{excess}} = \xi_0\mu/(2-\mu)$, since, for $\mu = 1$ and $\xi_0 \equiv \sigma_v^2 = 10^{-3}$, we have $\xi_0\mu/(2-\mu) = 10^{-3} \equiv -30$ dB.

Next, we tried to validate the claims made in Corollary 2 for which we considered the PNLMS algorithm and a white input while keeping all other parameters and the unknown system same as above. For each tap, the MSD $E[\tilde{w}_{P,i}^2(n_s)]$ was evaluated by averaging $\tilde{w}_{P,i}^2(n_s)$ over 1000 experiments, where n_s corresponds to a steady-state time index. For the active taps, the MSDs are plotted in Fig. 4, where, for comparison purpose, we also show the theoretical MSD value as per (5) by "+" and the MSD as per the formula presented in [12] by "◊." In addition, for clarity, as a representative case, we zoom the

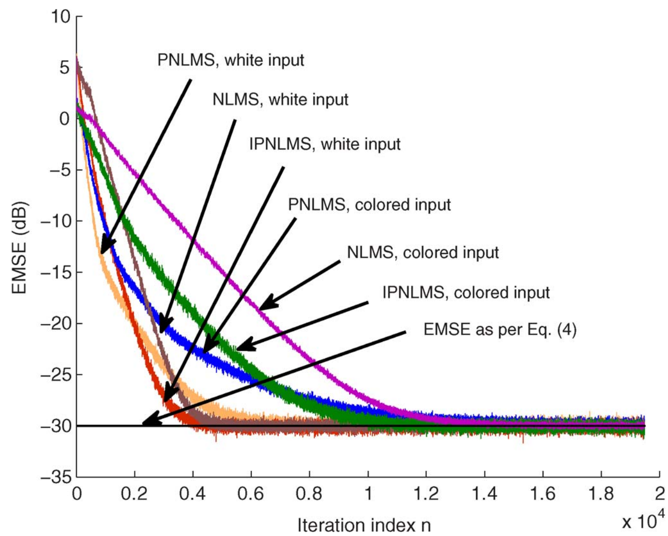

Fig. 3. Learning curves of the PNLMS, IPNLMS, and NLMS algorithms for white and colored inputs.

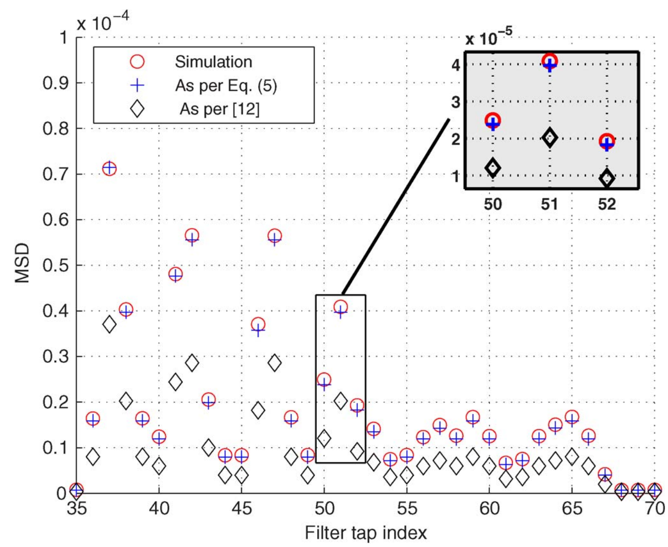

Fig. 4. Comparison of the steady-state MSD of the active taps as per simulation vis-a-vis as per the proposed analysis (i.e., (5)) and as per the expression derived earlier in [12].

region from the 50th tap to the 52nd tap. It is clearly seen that, for each tap, the simulated MSD and the MSD as per (5) are almost identical to each other but are much different from the MSD value obtained as per [12]. Same can be concluded from Fig. 5 as well, where the corresponding results for the inactive taps are shown separately. This validates the claims made in Corollary 2, which, in turn, confirms that Corollary 2 is a refinement of the existing result [12].

IV. Conclusion

A new convergence analysis has been presented for a well-known sparse adaptive filter family, namely, the PtNLMS algorithms, which evaluates the steady-state EMSE and conditions for stability of the EMSE. However, unlike the existing treatments in the literature, it does not assume the input to be white. Instead, it exploits a transform domain model of the PtNLMS

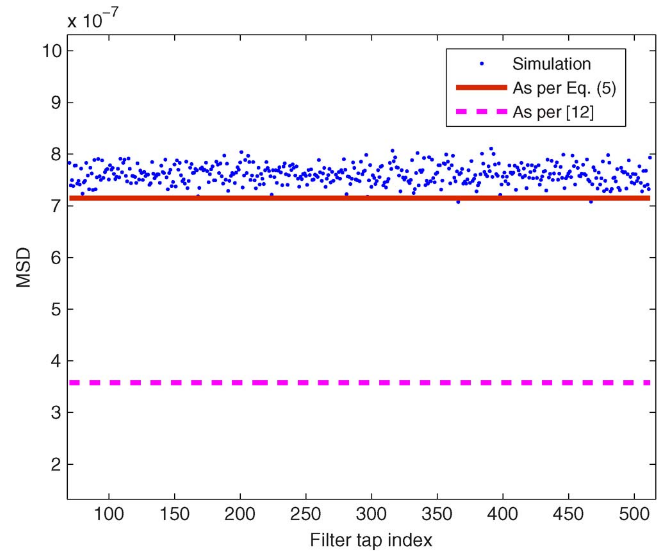

Fig. 5. Similar comparison of the steady-state MSD as in Fig. 4 applied to the inactive taps.

algorithm, in conjunction with a powerful scheme of angular discretization of a random vector. The proposed analysis reveals several new convergence features of the PtNLMS algorithms not reported earlier.

Appendix A
Proof of Theorem 1

Recalling that $\tilde{\mathbf{w}}_N(n+1) = \mathbf{G}^{-(1/2)}(n+1)\mathbf{w}_{\rm opt} - \mathbf{w}_N(n+1)$ and expressing $\mathbf{w}_N(n+1) \approx \mathbf{w}'_N(n+1)$ in terms of $\mathbf{w}_N(n)$ via the NLMS update equation, we have

$$\tilde{\mathbf{w}}_N(n+1) \approx \tilde{\mathbf{w}}_N(n) - \frac{\mu \mathbf{s}(n)\mathbf{s}^T(n)\tilde{\mathbf{w}}_N(n)}{\mathbf{s}^T(n)\mathbf{s}(n)+\delta} - \frac{\mu \mathbf{s}(n)v(n)}{\mathbf{s}^T(n)\mathbf{s}(n)+\delta} \tag{A.1}$$

where, as before, we assume that the diagonal entries $[\mathbf{G}(n)]_{i,i}$, $i = 0, 1, \ldots, L-1$ do not change significantly from the index n to the index $(n+1)$ [particularly near convergence], and thus, one can approximate $\mathbf{G}^{-(1/2)}(n+1)\mathbf{w}_{\rm opt}$ by $\mathbf{G}^{-(1/2)}(n)\mathbf{w}_{\rm opt}$. Using the "independence assumption" on $\mathbf{w}_N(n)$ vis-a-vis $\mathbf{s}(n)$, the recursive form of $\mathbf{K}_s(n)$ can be then written from (A.1) as

$$\begin{aligned}\mathbf{K}_s(n+1) &= E\left(\left(\mathbf{I} - \frac{\mu \mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\mathbf{K}_s(n)\left(\mathbf{I} - \frac{\mu \mathbf{s}(n)\mathbf{s}^T(n)}{\mathbf{s}^T(n)\mathbf{s}(n)}\right)\right) \\ &+ \mu^2 \xi_0 E\left(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\|\mathbf{s}(n)\|^4}\right)\end{aligned} \tag{A.2}$$

where we have ignored the cross-terms that become zero as $v(n)$ is zero mean and statistically independent of $\mathbf{s}(n)$. To simplify the last term, i.e., $E\left(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\|\mathbf{s}(n)\|^4}\right)$, we now invoke the angular discretization model of a random vector as discussed earlier and replace $\mathbf{s}(n)$ by $s_s(n)r_s(n)\mathbf{v}_s(n)$

as given by (3). One can then write $E\left(\frac{\mathbf{s}(n)\mathbf{s}^T(n)}{\|\mathbf{s}(n)\|^4}\right) = E\left(\frac{s_s^2(n)r_s^2(n)\mathbf{v}_s(n)\mathbf{v}_s^T(n)}{(s_s^2(n)r_s^2(n)\mathbf{v}_s^T(n)\mathbf{v}_s(n))^2}\right) = E\left(\frac{1}{r_s^2(n)}\right)\frac{\mathbf{S}(n)}{\mathrm{Tr}(\mathbf{S}(n))}$, since $s_s^2(n)=1$ and $\mathbf{S}(n)=\sum_{i=0}^{L-1}\lambda_{s,i}(n)\mathbf{e}_{s,i}(n)\mathbf{e}_{s,i}^T(n)$. Substituting into (A.2) and then premultiplying and postmultiplying $\mathbf{K}_s(n+1)$ by $\mathbf{e}_{s,i}^T(n+1)$ and $\mathbf{e}_{s,i}(n+1)$ respectively, we get

$$\begin{aligned}
&\tilde{\lambda}_{s,i}(n+1)\\
&\quad = \left(\mathbf{e}_{s,i}^T(n+1)\mathbf{K}_s(n+1)\mathbf{e}_{s,i}(n+1)\right)\\
&\quad = \sum_{j=1}^{L}\frac{\lambda_{s,j}(n)}{\mathrm{Tr}\,(\mathbf{S}(n))}\left(\mathbf{e}_{s,i}^T(n+1)\left(\mathbf{I}-\mu\mathbf{e}_{s,j}(n)\mathbf{e}_{s,j}^T(n)\right)\mathbf{K}_s(n)\right.\\
&\qquad\qquad \left.\times\left(\mathbf{I}-\mu\mathbf{e}_{s,j}(n)\mathbf{e}_{s,j}^T(n)\right)\mathbf{e}_{s,i}(n+1)\right)\\
&\qquad + \mu^2\xi_0 E\left(\frac{1}{r_s^2(n)}\right)\frac{\lambda_{s,i}(n)}{\mathrm{Tr}\,(\mathbf{S}(n))}.
\end{aligned}\tag{A.3}$$

As $\mathbf{G}(n)$ changes with time slowly (particularly near convergence), it will be quite reasonable to assume that the orthonormality relation between the ith and jth eigenvectors at the nth index, i.e., $\mathbf{e}_{s,i}^T(n)\mathbf{e}_{s,j}(n)=\delta(i-j)$, will continue to remain valid, albeit approximately, if $\mathbf{e}_{s,i}(n)$ is replaced by $\mathbf{e}_{s,i}(n+1)$ [note that, for white $x(n)$, this assumption is automatically satisfied, as in this case, $\mathbf{S}(n)$ $(=\sigma_x^2 E(\mathbf{G}(n)))$ is a diagonal matrix and thus has *time-invariant* eigenvectors given by the standard basis of $\mathbb{R}^L$]. This implies that the two orthonormal bases for $\mathbb{R}^L$, namely, $\{\mathbf{e}_{s,i}(n+1)|i=0,1,\ldots,L-1\}$ and $\{\mathbf{e}_{s,i}(n)|i=0,1,\ldots,L-1\}$, are closely aligned to each other. This permits the following approximations: $\mathbf{e}_{s,i}^T(n+1)\mathbf{K}_s(n)\times\mathbf{e}_{s,i}(n+1)\approx\mathbf{e}_{s,i}^T(n)\mathbf{K}_s(n)\mathbf{e}_{s,i}(n+1)\approx\mathbf{e}_{s,i}^T(n+1)\mathbf{K}_s(n)\times\mathbf{e}_{s,i}(n)\approx\mathbf{e}_{s,i}^T(n)\mathbf{K}_s(n)\mathbf{e}_{s,i}(n)=\tilde{\lambda}_{s,i}(n)$. Substituting into (A.3), we have

$$\begin{aligned}
\tilde{\lambda}_{s,i}(n+1) &= \frac{\lambda_{s,i}(n)}{\mathrm{Tr}\,(\mathbf{S}(n))}(1-\mu)^2\tilde{\lambda}_{s,i}(n)+\tilde{\lambda}_{s,i}(n)\\
&\quad\times\sum_{j\neq i}\frac{\lambda_{s,j}(n)}{\mathrm{Tr}\,(\mathbf{S}(n))}+\mu^2\xi_0 E\left(\frac{1}{r_s^2(n)}\right)\frac{\lambda_{s,i}(n)}{\mathrm{Tr}\,(\mathbf{S}(n))}\\
&= \left[1-\mu(2-\mu)\frac{\lambda_{s,i}(n)}{\mathrm{Tr}\,(\mathbf{S}(n))}\right]\tilde{\lambda}_{s,i}(n)\\
&\quad + \mu^2\xi_0 E\left(\frac{1}{r_s^2(n)}\right)\frac{\lambda_{s,i}(n)}{\mathrm{Tr}\,(\mathbf{S}(n))}.
\end{aligned}\tag{A.4}$$

From (A.4), convergence of $\tilde{\lambda}_{s,i}(n)$ requires $-1<1-\mu(2-\mu)\times\frac{\lambda_{s,i}(n)}{\mathrm{Tr}(\mathbf{S}(n))}<1$, which, after some calculations, leads to the following condition: $0<\mu<2$. Under this and recalling that $\mathrm{Tr}[\mathbf{S}(n)]=\sigma_x^2$, one then obtains the following from (A.4):

$$\tilde{\lambda}_{si}(\infty)=\frac{\xi_0\mu}{(2-\mu)}\lim_{n\to\infty}E\left(\frac{1}{r_s^2(n)}\right)=\frac{\xi_0\mu}{\hat{r}_s^2(\infty)(2-\mu)}.\tag{A.5}$$

Substituting (A.5) into (2) and recalling that $\sum_{i=0}^{L-1}\lambda_{s,i}(\infty)=Tr(\mathbf{S}(\infty))=\sigma_x^2$, it then easily follows that

$$\xi(\infty)=\xi_0+\frac{\xi_0\sigma_x^2\mu}{\hat{r}_s^2(\infty)(2-\mu)}.\tag{A.6}$$

References

[1] J. Radecki, Z. Zilic, and K. Radecka, "Echo cancellation in IP networks," in *Proc. 45th Midwest Symp. Circuits Syst.*, Tulsa, OK, USA, Aug. 2002, vol. 2, pp. 219–222.

[2] E. Hansler and G. Schmidt, Eds., *Topics in Acoustic Echo and Noise Control*. Berlin, Germany: Springer-Verlag, 2006.

[3] W. Schreiber, "Advanced television systems for terrestrial broadcasting," *Proc. IEEE*, vol. 83, no. 6, pp. 958–981, Jun. 1995.

[4] W. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Prof. IEEE CISS*, Princeton, NJ, USA, Mar. 2008, pp. 5–10.

[5] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proc. IEEE OCEANS*, 1995, vol. 3, pp. 1417–1422.

[6] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1986.

[7] R. L. Das and M. Chakraborty, "Sparse adaptive filters—An overview and some new results," in *Proc. ISCAS*, Seoul, South Korea, May 2012, pp. 2745–2748.

[8] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[9] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Orlando, FL, USA, 2002, pp. 1881–1884.

[10] H. Deng and M. Doroslovacki, "Improving convergence of the PNLMS algorithm for sparse impulse response identification," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 181–184, Mar. 2005.

[11] K. Wagner and M. Doroslovacki, *Proportionate-type Normalized Least Mean Square Algorithms*. Hoboken, NJ, USA: Wiley-ISTE, 2013.

[12] J. A. Garcia and A. R. Figueiras-Vidal, "Adaptive combination of proportionate filters for sparse echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 6, pp. 1087–1098, Aug. 2009.

[13] M. Nekuii and M. Atarodi, "A fast converging algorithm for network echo cancelation," *IEEE Signal Process. Lett.*, vol. 11, no. 4, pp. 427–430, Apr. 2004.

[14] K. Wagner and M. Doroslovacki, "Proportionate-type normalized least mean square algorithms with gain allocation motivated by mean-square-error minimization for white input," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2410–2415, May 2011.

[15] M. Yukawa and I. Yamada, "A unified view of adaptive variable-metric projection algorithms," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–13, 2009.

[16] Z. Yang, Y. R. Zheng, and S. L. Grant, "Proportionate affine projection sign algorithms for network echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 8, pp. 2273–2284, Nov. 2011.

[17] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York, NY, USA, Wiley, 2003.

[18] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2811–2825, Sep. 1993.

[19] T. K. Paul and T. Ogunfunmi, "On the convergence behavior of the affine projection algorithm for adaptive filters," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 8, pp. 1813–1826, Aug. 2011.